Claims

What is claimed is:

1. A computer-based method of performing document retrieval in accordance with an information network, the method comprising the steps of:

retrieving one or more documents from the information network that satisfy a user-defined predicate;

collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed; and

using the collected statistical information to automatically determine further document retrieval operations.

2. The method of claim 1, wherein the user-defined predicate specifies content associated with a document.

3. The method of claim 1, wherein the statistical information collection step uses content of the one or more retrieved documents.

4. The method of claim 1, wherein the statistical information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents.

5. The method of claim 1, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate.

6. The method of claim 1, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also satisfy the predicate.

7. The method of claim 1, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate.

8. The method of claim 1, wherein the information network is the world wide web and a document is a web page.

9. The method of claim 8, wherein the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.

10. Apparatus for performing document retrieval in accordance with an information network, the apparatus comprising:

at least one processor operative to: (i) retrieve one or more documents from the information network that satisfy a user-defined predicate; (ii) collect statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed; and (iii) use the collected statistical information to automatically determine further document retrieval operations.

11. The apparatus of claim 10, wherein the user-defined predicate specifies content associated with a document.

12. The apparatus of claim 10, wherein the statistical information collection operation uses content of the one or more retrieved documents.

13. The apparatus of claim 10, wherein the statistical information collection operation considers whether the user-defined predicate has been satisfied by the one or more retrieved documents.

14. The apparatus of claim 10, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate.

15. The apparatus of claim 10, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also satisfy the predicate.

16. The apparatus of claim 10, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate.

17. The apparatus of claim 10, wherein the information network is the world wide web and a document is a web page.

18. The apparatus of claim 17, wherein the statistical information collection operation uses one or more uniform resource locator tokens in the one or more retrieved web pages.

19. An article of manufacture for performing document retrieval in accordance with an information network, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

retrieving one or more documents from the information network that satisfy a user-defined predicate;

collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed; and

using the collected statistical information to automatically determine further document retrieval operations.

20. The article of claim 19, wherein the user-defined predicate specifies content associated with a document.

5   21. The article of claim 19, wherein the statistical information collection step uses content of the one or more retrieved documents.

22. The article of claim 19, wherein the statistical information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents.

10   23. The article of claim 19, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate.

24. The article of claim 19, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the

15   one or more retrieved documents that also satisfy the predicate.

25. The article of claim 19, wherein the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate.

26. The article of claim 19, wherein the information network is the world wide

20   web and a document is a web page.

27. The article of claim 26, wherein the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.